FUSED VISUALIZATION OF COMPLEX INFORMATION SPACES

By

Wu Quan

Submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Technology, Sydney

July 2006

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

To My Wife, Yu Liu

Acknowledgements

I would like to express my deep gratitude to my supervisor, Dr. Mao Lin Huang, for providing me with academic guidance, inspiration and motivation throughout the course of this research work. Without his care, supervision and friendship, I would not be able to complete my work. Meanwhile, I am impressed by his enthusiasm, persistence and diligence in academic research, which will benefit my future career and life.

My thanks also go to A/Prof. Tom Hintz and Prof. Chengqi Zhang, for helping me get the opportunity to pursue doctorate at the Faculty of Information Technology at the University of Technology, Sydney.

I would also like to appreciate Dr. Quang Vinh Nguyen, Dr. Wu Qiang, A/Prof. Xiangjian He and all the other members of Computer Vision Research Group for their generous helps and suggestions in my research work.

I would like to express my thanks to the Faculty of Information Technology for offering me a Research Scholarship throughout my doctoral program, and for providing me with the financial support for attending several conferences.

Last but not least, I am indebted to my wife Yu Liu who accompanies me and takes care of my life during my study. Without her care, patience and continuous encouragement, I would not be able to finish my study. Great thanks are also due to my parents, parents-in-law, my elder sister Wen Quan, brother-in-law Yuanqing Liu and my cousin Lan Quan, Xiaohui Zhu, Zhiqiang Zhu for their forever affection and constant encouragement.

Contents

List of Figures

Terminology

Several terms are used throughout this thesis. They are described below:

Analytical Information Visualization: is a visualization that seeks ways to support the understanding of analyzed information through the use of appropriate visual metaphors, and corresponding geometric layouts and navigation methods.

Visual Metaphor: is a visual representation of a data set by means of visual attributes corresponding to a different system that behaves in a similar way and is well-known to users.

A Graph *G(V, E)*: is defined as a pair *(V, E)* consisting of a finite set *V* of vertices and a finite set *E* of edges , where $E = \{(u,v) \mid u,v \in V\}$.

A Visualization Method V_m: could be a modeling scheme, a layout algorithm, or a viewing technique etc.

A set of Properties $P(V_m)$: is a set of technical features of V_m. These technical features could be high interactive speed, high efficiency of space utilization, low computational complexity, different physical references and so on.

A Focus Frame $F_i(G_i, Q_i)$: consists of a connected sub-graph $G_i = (V_i, E_i)$ of G and a queue Q_i of 'focus' vertices. Successive focus frames differ only in a few vertices. This sequence of focus frames is the sequence of sub-graphs of G viewed by the user.

A Spatial Frame S_f: is an interactive thematic map that consists of a set of spatial regions. Thus, we have $S_f = \{R_1, R_2, \dots, R_n\}$. There is an independent graph G_i can be derived from a corresponding spatial region R_i.

A Spatial Region *R*: is a district in the spatial frame (such as a state of a country on the map). We define each spatial region as $R_i = (name_i, x_i, y_i, w_i, h_i, polygon_i, S_f^{R_i})$, in which the $name_i$ is the identifier of the region, x_i *and* y_i are coordinators of the top-left point of the rectangle that just bounds the polygonal boundary of R_i, w_i and h_i is the width and height of the rectangle respectively, $polygon_i$ contains a series of the boundary points of the region, S_f^{Ri} is a lower level nested spatial frame that contains the sub-regions information of R_i.

A Circular Region *C(G)*: is a display space that contains the drawing of a graph G_i. The geometric position of a $C(G_i)$ in the visualization depends on the position of R_i from which the graph G_i is derived.

A Historic Trail H_t: consists of a sequence of small circles displayed above the spatial frame, which represents the recently exploring history of graphs.

A Drawing $D(G_i)$: is a geometric drawing of graph G_i. It consists of a position for each vertex $v \in V_i$ and a route for each edge $e \in E_i$.

Time Consumption *TC*: is the running time spent in drawing a graph from its initial state to the end state.

Vertex Position Variance *VPV*: indicates the position change of a vertex from its start position to its end position in the drawing.

Vertex Travel Length *VTL*: is the length of the entire path that a vertex travels from its start position to its end position of the layout, which may consists of a series of turning points P_s, P_2, ...P_{j-1}, P_j, ..., P_e.

Abstract

With the rapid growth of information analysis and data mining technologies, the massive data sets available for access have been merged and refined to manifold information, including raw data and all kinds of analytical results. Since data sets become increasingly complex, the current visual analytical techniques no longer satisfy the needs of exploring and analyzing data.

This situation raises the challenges in the current state of information visualization:

1) Due to the complexity of information, sometimes it is unlikely to use a single visual metaphor to model the intricate information well in a single visualization.

2) Each existing visualization method has its own limitations in terms of satisfying domain specific requirements, when dealing with complex data sets.

The proposed fused visualization methodology attempts to address the above issues by combining multiple existing visualization techniques in a single visualization. It takes the advantages and reduces the weaknesses of the existing methods. We have successfully applied this methodology to each stage of the proposed *Analytical Information Visualization*.

In particular, three fused visualization techniques are developed to improve the quality of existing techniques. First, a fused visual metaphor that combines two visual metaphors in a single visualization allows users to navigate spatially referenced information across two different metaphors. Second, a fused layout algorithm that combines two graph drawing methods achieves the fast convergence in geometric layout for the force-directed layout algorithm; Third, a fused viewing technique that combines 1D and 2D distortional visual viewing methods in one browser resolves the inefficient space utilization problem.

Moreover, the fused layout algorithm has been evaluated against other existing force-directed layout algorithms. Two case studies that apply our techniques to an *outbreak management system* and an *online bookstore* respectively have been delivered.